Université de Versailles – Saint-Quentin-en-Yvelines
Université Paris Saclay

le _ _ / _ _ / 2024

Master 1 Mathématiques 2023–2024
Théorie de l'Information

NOM : ____________________	Prénom : __________________	Num. Étu. : 2 ☐☐☐☐☐☐☐

Questions :

1. Soient X et Y deux v.a. Donner la définition mathématique de l'information mutuelle moyenne, de l'entropie de X sachant Y et de l'entropie de X, puis donner et montrer la relation entre ces trois quantités.
2. Soient X, Y, Z trois v.a. dans $\{0, 1\}$ avec

$$p_{XYZ}(0,0,0) = \frac{1}{4} = p_{XYZ}(1,0,0) = p_{XYZ}(0,1,0) = p_{XYZ}(1,0,1)$$

   Calculez $H(X)$, $H(Y|X)$, $H(Z|X,Y)$. En déduire $H(X,Y,Z)$. Calculez $H(Y)$. Combien d'information apporte X sur Y et réciproquement ?
3. Une pièce est lancée jusqu'à l'occurence d'une face où la probabilité d'avoir face est p. On considère X la v.a du nombre de lancements. Déterminez $H(X)$. On considère maintenant Y la v.a. correspondant au nombre de lancers jusqu'à avoir 2 faces. Justifiez précisément pourquoi $H(Y) < 2H(X)$.

Réponses :

Question 1 X et Y sont deux v.a. à valeurs dans $\mathcal{X}$ et $\mathcal{Y}$ respectivement munies de lois de probabilités notées $p_X(x)$ pour $x \in \mathcal{X}$ et $p_Y(y)$ pour $y \in \mathcal{Y}$. On note $p_{XY}(x,y) = P[X = x, Y = y]$ la loi de probabilité jointe. On rappelle aussi que l'on a les relations suivantes pour tout $x \in \mathcal{X}$ et pout tout $y \in \mathcal{Y}$.
— $p_{XY}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)}$ et $p_{XY}(y|x) = \frac{p_{XY}(x,y)}{p_X(x)}$
— $p_X(x) = \sum_{y\in\mathcal{Y}} p_{XY}(x,y)$ et $p_Y(y) = \sum_{x\in\mathcal{X}} p_{XY}(x,y)$
L'information mutuelle moyenne notée $I(X;Y)$ entre X et Y est donnée par la formule

$$I(X;Y) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p_{XY}(x,y) \log_2\left(\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}\right).$$

L'entropie de X sachant Y notée $H(X|Y)$ est la moyenne sur tout l'espace probabilisé $\mathcal{X} \times \mathcal{Y}$ de l'information propre des événements $X = x|Y = y$, c'est à dire

$$H(X|Y) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} -p_{XY}(x,y) \log_2(p_{XY}(x|y)).$$

L'entropie de X notée $H(X)$ est la moyenne de l'information propre, c'est à dire

$$H(X) = \sum_{x\in\mathcal{X}} -p_X(x) \log_2(p_X(x)).$$

On sait que $I(X;Y) = H(X) - H(X|Y)$. En effet,

$$I(X;Y) + H(X|Y) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p_{XY}(x,y) \log_2\left(\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}\right) - p_{XY}(x,y) \log_2(p_{XY}(x|y))$$

Donc
$$I(X;Y) + H(X|Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{XY}(x,y)\log_2\left(\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)p_{XY}(x|y)}\right)$$
Donc
$$I(X;Y) + H(X|Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{XY}(x,y)\log_2\left(\frac{1}{p_X(x)}\right) = \sum_{x\in\mathcal{X}}\log_2\left(\frac{1}{p_X(x)}\right)\left(\sum_{y\in\mathcal{Y}} p_{XY}(x,y)\right)$$
Ce qui est égal à
$$\sum_{x\in\mathcal{X}} -p_X(x)\log_2(p_X(x)) = H(X)\,.$$

Question 2 Pour calculer $H(X)$ on a besoin de la loi marginale de X qui est obtenue en réalisant
$$p_X(0) = \sum_{(y,z)\in\{0,1\}^2} p_{XYZ}(0,y,z) = \frac{1}{4}+\frac{1}{4} = \frac{1}{2} = 1 - p_X(1) = p_X(1)\,.$$
La valeur de H vient immédiatement car on est sur une loi uniforme sur un espace de taille 2 ce qui équivaut à la définition du bit d'information : l'entropie $H(X)$ vaut 1 bit.
$$H(Y|X) = \sum_{x,y} p_{XY}(x,y)\log_2(p_{XY}(y|x))\,.$$
Pour calculer $p_{XY}(x,y)$, on réalise la somme $\sum_z p_{XYZ}(x,y,z)$ pour toute valeur x,y. On obtient ici $p_{XY}(0,0) = \frac{1}{4} = p_{XY}(0,1)$ et $p_{XY}(1,0) = \frac{1}{2}$ et enfin $p_{XY}(1,1) = 0$. Ensuite, on cherche $p(y|x)$ qui est égal à $\frac{p_{XY}(x,y)}{p_X(x)}$. On a déjà calculé la loi marginale $p_X(x)$ qui est uniforme. Alors,
$$H(Y|X) = -\frac{1}{4}\log_2\left(\frac{\frac{1}{4}}{\frac{1}{2}}\right) - \frac{1}{2}\log_2\left(\frac{\frac{1}{2}}{\frac{1}{2}}\right) - \frac{1}{4}\log_2\left(\frac{\frac{1}{4}}{\frac{1}{2}}\right) = \frac{1}{2}\,.$$
Pour l'entropie $H(Z|X,Y)$ on fait la même chose. On regarde les 4 cas possibles et on obtient que $p(0|0,0) = 1 = p(1|0,1)$, $p(0|1,0) = \frac{1}{2} = p(1|1,0)$. Toutes les autres probabilités valent 0. On obtient alors que
$$H(Z|X,Y) = -2\frac{1}{4}\log_2\left(\frac{1}{2}\right) = \frac{1}{2}\,.$$
D'après la règle de chainage, on sait que
$$H(X,Y,Z) = H(Z|X,Y) + H(Y|X) + H(X) = 2\,.$$
Ceci est tout à fait normal et aurait pu être trouvé directement : les variables X, Y, Z peuvent être aussi vues comme une seule variable, avec 4 possibilités équiprobable, et donc l'entropie et le logarithme de la taille de l'espace, ici 2.

Pour calculer $H(Y)$, on doit maintenant calculer les lois marginales, c'est à dire $p_Y(y)$ et on a
$$p_Y(0) = \sum_{x,z} p_{XYZ}(x,0,z) = \frac{3}{4} = 1 - p_Y(1)\,.$$
Ainsi,
$$H(Y) = \frac{3}{4}\log_2\left(\frac{4}{3}\right) + \frac{1}{4}\log_2(4) = 2 - \frac{3}{4}\log_2(3)$$
Ce qui est nécessairement positif (car l'entropie est positive) et inférieur strictement à 1 (car la loi n'est pas uniforme sur un ensemble de taille 2).

Enfin pour calculer combien apporte d'information X sur Y et respectivement, on calcule l'information mutuelle $I(X;Y) = H(X) - H(X|Y) = \frac{1}{2}$. Seule, cette information ne donne pas grand chose. Il faut la comparer à $H(X)$ et $H(Y)$. Comme $H(X) > H(Y)$, Y apporte plus d'information sur X que X sur Y car Y car on a moins d'incertitude a priori sur Y que sur X.

Question 3 La loi de probabilité de X est donnée par la formule

$$p_X(x) = p(1-p)^{x-1}$$

En effet, il faut réaliser exactement $x-1$ tirages piles puis un tirage face pour s'arrêter. Cela nous dit donc que

$$H(X) = \sum_{x=1}^{\infty} p(1-p)^{x-1} \log_2 \left(\frac{1}{p(1-p)^{x-1}} \right) = -p \log_2(p) \left(\sum_{n=0}^{\infty} (1-p)^n \right) - p \log_2(1-p) \left(\sum_{n=0}^{\infty} n(1-p)^n \right)$$

Ce qui peut alors s'écrire comme

$$H(X) = -p \log_2(p) \frac{1}{p} - p \log_2(1-p) \frac{1-p}{p^2} = \frac{H(p)}{p}$$

où $H(p)$ est l'entropie d'une v.a dans un espace à 2 éléments avec une distribution de probabilité $(p, 1-p)$.

Pour la deuxième partie de la question, on ne fera pas de calculs, mais on considère Y la v.a qui correspond à s'arrêter après avoir trouvé deux faces. Y peut être vue comme la somme de deux v.a X décrites précédemment que l'on notera X_1 et X_2. On peut directement dire que X_1 et X_2 sont indépendantes (car les lancés de pièces sont indépendants). On considère alors

$$H(X_1, X_2) = H(X_1) + H(X_2) = 2H(X)$$

(car les v.a sont indépendantes). On s'intéresse maintenant à $H(Y)$ et on utilise la règle de chainage. Ainsi,

$$H(Y) = H(Y, X_1) - H(X_1|Y)$$

Or si on "connaît" entièrement Y ou sa réalisation, il est clair qu'on ne détermine pas la valeur de X_1 (seulement une borne supérieure). Dit autrement, $H(X_1|Y) > 0$. Donc $H(Y) < H(Y, X_1)$. En utilisant encore la règle de chainage, on a aussi que

$$H(Y, X_1) = H(X_1) + H(Y|X_1) = H(X_1) + H(X_1 + X_2|X_1).$$

Mais, à X_1 connu, l'incertitude sur $X_1 + X_2$ ne réside que dans la v.a X_2, donc $H(X_1 + X_2|X_1) = H(X_2|X_1)$. Comme les v.a X_1 et X_2 sont indépendantes, on obtient que $H(Y, X_1) = 2H(X)$.